

# Quo vadis, EAD?

**Current use cases - from single institutions to aggregators and providers of archival management systems**

2021-06-02

## **Presenters**

Karin Bredenberg, co-chair TS EAS


Ana Maria López Cuadrado, Archivo Histórico Nacional

Jane Stevenson, Jisc

Christine di Bella, ArchivesSpace

# TS-EAS

- Technical Subcommittee on Encoded Archival Standards at the Society of American Archivists (SAA)
- Background and work covered in this presentation:
  - <https://youtu.be/9NXNyx9py-l>



We take care of the  
formats you use to  
manage and share  
archival information

We need your  
comments,  
suggestions and  
bug reports

Your input  
drives our  
work forward

# Where you can find more

TS EAS

<https://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-standards-ts-eas>



TS-EAS on GitHub

<https://github.com/SAA-SDT>



EAD publication

<http://www.loc.gov/ead/index.html>



EAC publication

<https://eac.staatsbibliothek-berlin.de/>



Our mailing list

[EAD@LISTSERV.LOC.GOV](mailto:EAD@LISTSERV.LOC.GOV)

Reporting an issue via SAA

<https://www2.archivists.org/standards/TS-EAS-report-an-issue>



# Standards revision

Annual rolling revision cycle for minor releases

(see more on GitHub:

<https://github.com/SAA-SDT/TS-EAS-subteam-notes/blob/master/rolling-revision-cycle.md>)

Evaluate standards for potential major revision every five years following guidelines by SAA's Standards Committee

(see more:

<https://www2.archivists.org/governance/handbook/section7/groups/Standards/Development-and-Review> )

# **Presentations**

---

# **PARES 2.0.**

# **EAD/EAC CPF/EAG 2012**

**ANA LÓPEZ CUADRADO**

**NATIONAL HISTORICAL ARCHIVE OF SPAIN**



# PARES 2.0.

- The [Portal of Spanish Archives \(PARES\)](#) is the main platform for the dissemination of Spanish Historical Documentary Heritage, created and managed by the General Subdirectorate of the State Archives Culture and Sports Ministry.
- PARES offers free and open access, both to researchers and any citizen interested in accessing to the digitized collections of the Spanish Archives.

# PARES 2.0.

- PARES 2.0. allow us to descriptive all type of the entities defined by NEDA (Spanish Archival Conceptual Model) and provides an efficient and effective access to information in a fast and intuitive way.
- Furthermore, PARES shows digitized images of the documentary collections preserved in the Spanish State Archives.



# PARES 2.0.

- PARES started to work in 2007. In 2018 evolved to PARES 2.0.
- The portal was conceived as a relational database, but it has been improved thanks to the application of the Spanish Conceptual Model. It was created to describe records, but it has become a tool that describes all kinds of entities that are related to each other and linked to external data.

# PARES 2.0.

- Since 2018, the PARES development has taken a leap: Now it captures data automatically through APIs from other cultural institutions and from the different technologies developed around the semantic web.
- During the description of records, archivists study them (as a primary source) while taking into account their context. All entities that are represented in the record are relevant and must be related to each other and to other records.

# PARES 2.0.

- Thus, our records system grows by a considerable number of related entities and links between them. This enables us to create a real navigation effect.
- These interconnections will increase exponentially if we not only reuse the information but also become content providers, changing our tool into a Linked Open Data-system.
- PARES 2.0. proves Records in Context or the multi-entity description functions in practice.

# PARES 2.0.

- On the other hand, PARES also allows us to export our information due to the great work of standardization that has been carried out over the years.
- PARES interoperates with other archival institutions, such as the European Archives Portal through the EAD, EAC CPF and EAG 2012 standards.
- It should be noted that already in 2002 Spain launched the first encoded archival standard for institutions with archival holdings, called EAG 2002. This standard was later revised and updated by the European Archives Portal, which currently manages and supports it.

# EAD / PARES 2.0.

- Focusing on our experience using EAD through PARES, we can take stock of the pros and cons of this standard.
- On the one hand EAD is a solid standard, and fully consolidated in the international panorama as a standard for the exchange of archival information.

# EAD / PARES 2.0.

- Our system is prepared to export information in EAD and EAD3. We mainly did the adjustment to implement the use of additional finding aids. Although we have implemented this possibility, our conceptual model is more focused on the use of document entities to describe the context and we don't find this development useful.
- Therefore, for us the adaptation to EAD3 has consisted of a conversion, revision and update of tags.



# EAD / PARES 2.0.

- We could talk at greater length about the benefits that the use of EAD has brought us, not so much about importing data as exporting information with various archival institutions.
- But considering the forum in which we find ourselves, I think it is more interesting to contribute ideas for improvement. We believe EAD should improve and evolve in the same way that new technologies and archival theory do.

# EAD IMPROVENTES

## Categorize external links

We found that the tags that allow us to include links, they don't let us to create enough categories to distinguish the types of external links.

There is an increasing tendency to use links that allow interconnectivity, so that the files present increasingly more external links that should be categorized.

If we categorize these links, EAD will gain in semantic structure.



# EAD IMPROVEMENTS

**Design a standard way of splitting the documents**

The nature of our hierarchical descriptions means that, at least in PARES 2.0, we get very large files.

APE has solved this problem with a split between Holding Guides and Finding Aids, creating a link between the two files, so that they connect seamlessly.

# EAD IMPROVEMENTS

Design a standard way of splitting the documents (2)

Even so, very large files continue to be produced, which sometimes make it difficult to automate the sending of information with systems such as OAI PMH.

There are options to make these partitions, but it would be very interesting to establish an official way to do it. In our case, this improvement would be very useful because this would mean be able to start the automation of data transmission

# EAD IMPROVEMENTS

## Tags which may include controlled vocabularies

Taking into account the developments proposed by the RIC, it would be very interesting if EAD allowed the definition of controlled vocabularies within the file itself.

If it were possible to include tags, they would allow the controlled vocabularies to be previously defined, being able to explain inside the EAD file what is an entity and what is not.

# EAD IMPROVENTES

**Tags which may include controlled vocabularies (2)**

In this way, a big step would be taken to relate and generate correspondence between EAD and RDF, bringing an eminently syntactic language closer to the semantics, something necessary in the RIC philosophy.

# THANKS!!! QUESTIONS?

ANA LÓPEZ CUADRADO

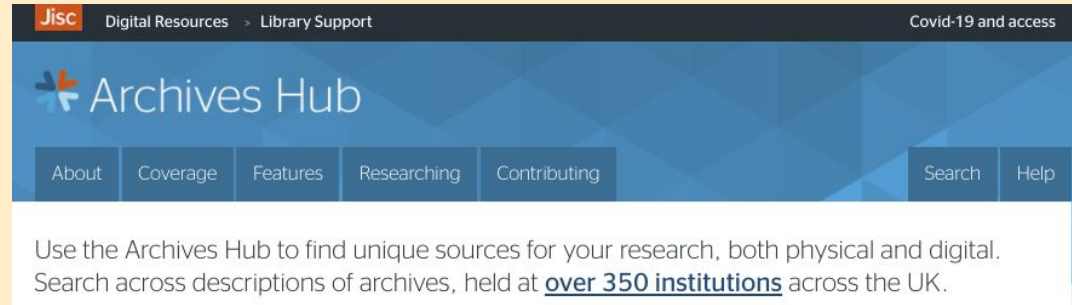
ANAMARIA.LOPEZ@CULTURA.GOB.ES

NATIONAL HISTORICAL ARCHIVE OF SPAIN



# Archives Hub

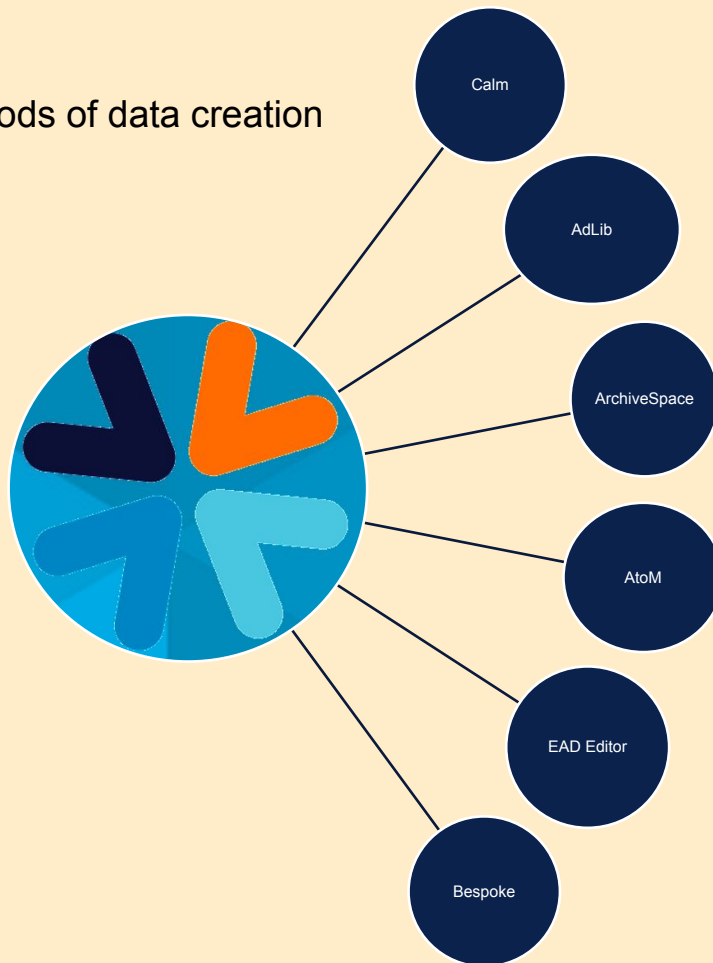
## and the use of EAD



*<https://archiveshub.jisc.ac.uk>*

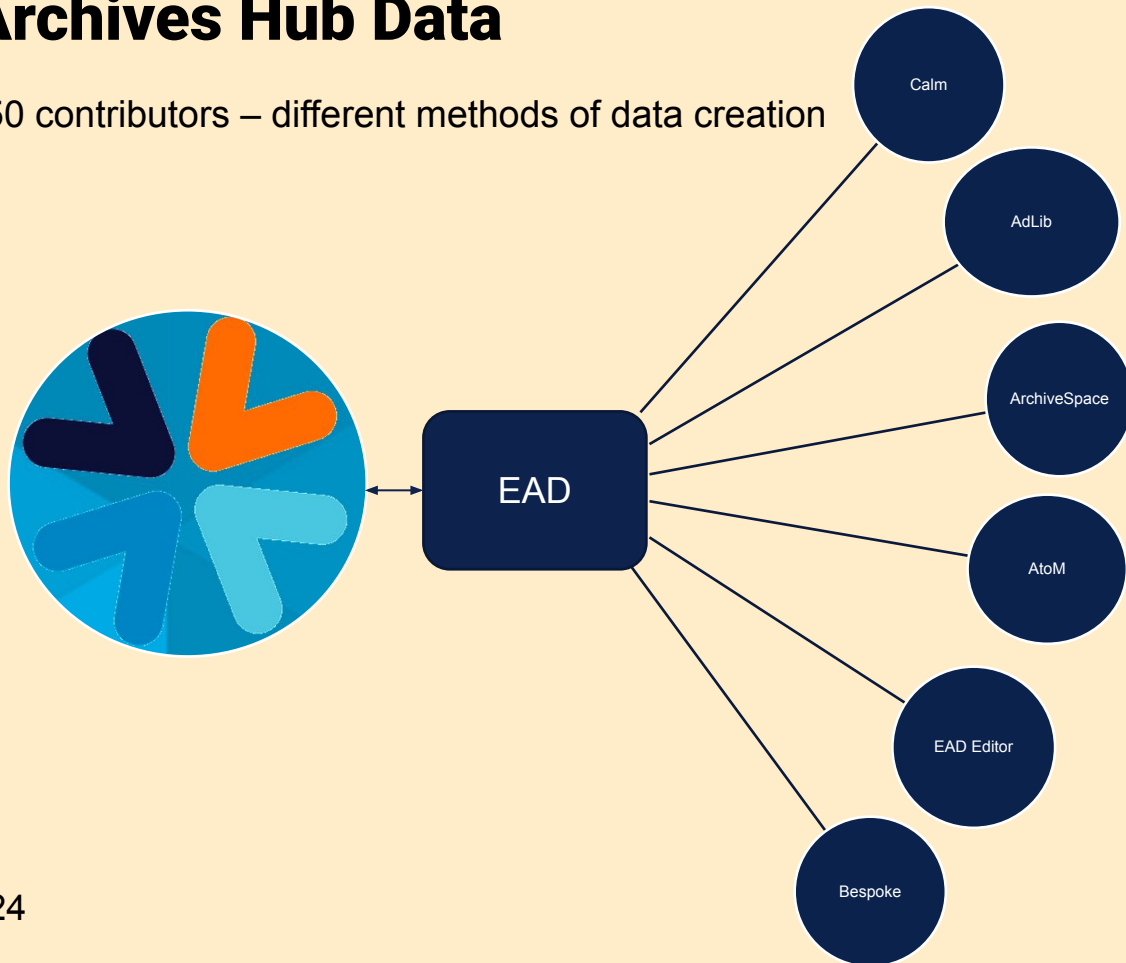
# Archives Hub Data

350 contributors – different methods of data creation



# Archives Hub Data

350 contributors – different methods of data creation





# How is EAD used?

- EAD for Data in / Data out
- Converted to JSON within the system (the CIIM)
- Data ingest validates against the EAD2002 DTD
- Data is 'normalised' via our ingest pipelines (Archives Hub validation)
- All descriptions can be downloaded – original and processed

## Eyles Collection

### Archive Collection

For more information, email the repository  
Advice on accessing these materials  
Cite this description  
Bookmark: <https://archiveshub.jisc.ac.uk/data/gb3-dm1186>

XML

This material is held at	University of Bristol Special Collections
Reference	GB 3 DM 1186
Dates of Creation	1679-1983
Name of Creator	Victor and Joan Eyles
Language of Material	English, fra deu
Physical Description	5 archive boxes

<input type="checkbox"/>		<b>EAD</b>	<b>Eyles Collection</b> DM 1186 bristol 0035_2002_2002.xml	Published <a href="#">View Audits</a>	<a href="#">Actions</a> ▼ Remove from Workflow Repeat Global Transforms Reprocess Original Send to ES Unpublish <b>Download:</b> original current processed
<input type="checkbox"/>		<b>EAD</b>	<b>Papers and correspondence of Sir Frederick Charles Frank, 1911-1996.</b> DM 1310 and DM 1837 bristol 59_2002.xml	Published <a href="#">View Audits</a>	
<input type="checkbox"/>		<b>EAD</b>	<b>Radstock Labour Party Papers</b> DM 1132 bristol 0057_2002_2002.xml	Published <a href="#">View Audits</a>	
<input type="checkbox"/>		<b>EAD</b>	<b>Leigh Manor Rolls</b> DM 177 bristol 0045_2002_2002.xml	Published <a href="#">View Audits</a>	

# Why EAD2002?

- Converted all our data from EAD v.1 in about 2001 - big job, lots of tidying up to do
- We see EAD as a data exchange format
- Cannot see a solid business case for moving to EAD3
  - physdesc/physdescstructured was a big problem
  - already have name structure markup
  - cost/benefit of EAD2002 to EAD3 vs. other developments
- Hardly any contributors use EAD themselves
- EAD we work with is exports from Archival Management Systems – no EAD3 exports

# Physical description

```
<physdesc>  
<extent>37 boxes</extent>  
<genreform>manuscripts, typescripts, photographs, ephemera</genreform>  
</physdesc>
```

```
<physdesc>  
<extent>1 volume</extent>  
<dimensions>18cm x 28cm</dimensions>  
<physfacet>blue leather binding with gold lettering</physfacet>  
</physdesc>
```

```
<physdescstructured coverage="whole"  
physdescstructuredtype="spaceoccupied">  
  <quantity>12</quantity> <unittype> cubic metres</unittype>  
</physdescstructured>
```

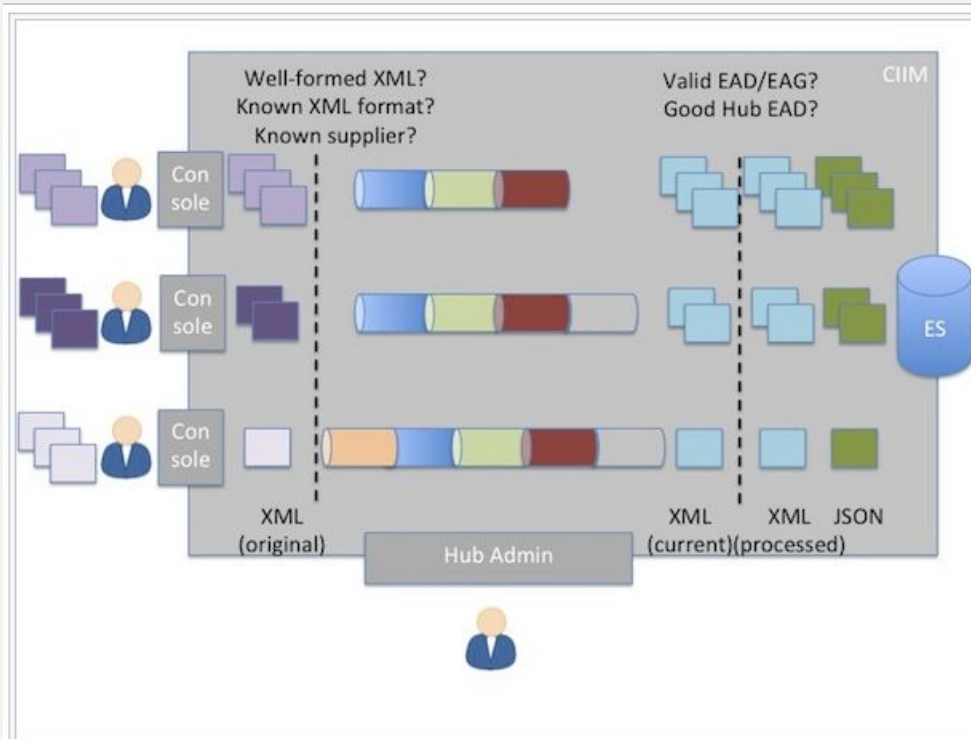
# Access Points

```
<persname source="viaf" rules="ncarules" authfilenumber="https://viaf.org/viaf/52493570">
  <emph altrender="surname">Smith</emph>
  <emph altrender="forename">John</emph>
  <emph altrender="dates">1938-1994</emph>
  <emph altrender="epithet">labour politician</emph>
</persname>
```

```
<persname rules="ncarules">
  <part localtype="surname">Smith</part>
  <part localtype="forename">John</part>
  <part localtype="dates">1938-1994</part>
  <part localtype="epithet">Labour politician</part>
</persname>
```

# **EAD as a language**

# Data processing



Data transformation within the CIIM

# XSLT processing

## TransformationXSL remove-audience-internal

### Contents [\[hide\]](#)

- 1 [What It Does](#)
- 2 [Dependencies](#)
- 3 [Parameters](#)
- 4 [Configuration](#)
- 5 [Run by Script](#)

### What It Does [\[edit\]](#)

- This is step 14 of the Global pipeline
- Removes elements with attribute audience='internal'
- Removes comments
- Removes processing instructions
- Removes most filedesc/publicationstmt data (because historically it often includes institutional contact data that becomes outdated)

# The language of EAD

## Analysis of Sample Data (20170201) [\[edit\]](#)

### Notes [\[edit\]](#)

- 11 docs
- unitid
  - unitid[@label='Reference'] (RefNo) 10001
  - unitid[@label='Former Reference'] (AltRefNo) 8973
- level
  - otherlevel='Group'
    - map to level='recordgrp'
- unitdate
  - unitdate 9987
  - unitdate/@normal null 281
  - content in format
    - dd-MMMM-yyyy
    - dd-MMMM-yy
    - yyyy-yyyy
      - few other variants
  - @normal is a range, years only, even for single dates
  - @normal null for yy=00, 01, 02 etc up to 10 (21st century)
  - @normal is bad (but 'valid') for yy=72, 73 etc (20th century)
- langmaterial
  - OK
- origination
  - present in 11 docs
  - 549 occurs
  - encodinganalog="Calm CreatorName"
- controlaccess
  - persname 0
  - corpname 0
  - famname 0
  - geogname 1
  - subject 753

## Brief analysis of Sample Data (Abercrombie - sent 20170406)

- eadheader: eadid hopeless; no titleproper
  - unitid: need repositorycode in order to process
  - no repository
  - no unitdate/@normal
  - top-level unittitle content often just a name e.g. Abercrombie, Charles
  - controlaccess need to check names and places do export
  - any fields not exporting?
- 20200120: JS sent an email re. exporting. They sent a sample, but it is a <c> entry (see Modes export folder). Hoping to clarify if the export only provides <c> entries, whether they catalogue at series level, and if not, whether we can get a full collection description. They are keen to progress, but it is a case of finding the time to do this.
  - 20200120: They sent a larger sample - we can use this to analyse the EAD and feedback to BAS and to Modes (AD series)
  - 20210211: Their langmaterial data in directly in the <c> tags so KC has written a script to move these into the <did> where the DTD expects them. This script will be run in a BAS pipeline before the Modes Group pipeline

## Data Sample from early 2019 [\[edit\]](#)

- Extent – required just at top level, so we do need this.
- langmaterial: the sample had: <langmaterial>English</langmaterial> - we would have to do some processing to make this correct EAD, and we'll have to see more samples with e.g. several languages, to ensure that we can process this OK.
- 'odd' displays twice because it includes 'note' and our processing displays anything in <odd> and anything in <note>, but I think we can look at doing something to stop this displaying twice, so not a big problem.



# Exporting from archival management systems

Access conditions	<accessrestrict>	To provide information on the legal status or other regulations that restrict or affect access to the unit of description	<p>This collection is open for research. It is advisable to book in advance, for details and opening hours see <a href="http://archives.denbighshire.gov.uk/visit-us/">http://archives.denbighshire.gov.uk/visit-us/</a>. Access to some documents may be restricted due to Data Protection legislation, Denbighshire Record Office will advise where this is the case</p>	<p>- Standard text can be included in a picklist e.g. 'Some records are subject to access restrictions under the Data Protection Act'</p> <p>-This field is mandatory for the Archives Hub at the top level because it is vital researchers know whether they can access the materials.</p> <p>*PICKLIST</p>
Language	<language>	To identify the language(s), used in the material being described	<p>Welsh English French</p>	<p>- If the material is in more than one language, they should be recorded in separate language fields commencing with the predominant language(s) (To add fields go to &lt;field&gt;, select &lt;insert&gt;, then choose 'Language')</p> <p>- The Archives Hub will add the ISO code for the language, e.g. French (fre)</p>

# EAD template

```
<!-- MINIMAL TEMPLATE with notes
This template provides for the minimal level of EAD necessary for the Archives Hub.
Our processing should be able to make the necessary changes to create Hub EAD.
The STANDARD TEMPLATE includes more structured data to create fully valid EAD.
-->
<!-- The preferred character encoding for the Hub is UTF-8. Other encoding can be used as long as
the data reflects the specified encoding -->
<!-- EAD DOCUMENT -->
<ead>
  <!-- A single eadheader element is required
  We can generate the basic header, to include title and reference -->
  <eadheader> </eadheader>
  <!-- DESCRIPTION -->
  <!-- A single archdesc element is required -->
  <!-- For the archdesc element
  - the level attribute is required with a value from the EAD controlled list
  - if the 'otherlevel' attribute is used, the value must be from the Hub controlled list
  -->
  <archdesc level="1">
    <!-- CORE INFORMATION - A single did element is required -->
    <did>
      <!-- For the reference we need the countrycode and repository code and local identifier
      to be provided
      We can remove leading zeros in repository codes
      We will generate an identifier attribute as long as the reference is clear
      The Hub supports former and alternative references as well. For information on these
      please see the STANDARD template -->
      <unitid countrycode="GB" repositorycode="123">MSC</unitid>
      <!-- A single unittitle element is required -->
      <unittitle/>
      <!-- For each date range, use a separate unitdate element -->
      <!-- Inclusion of the normal attribute allows for date searching/filtering - see the
      STANDARD template. We may be able to populate the normalised date if the date given follows defined
      formatting
      -->
      <unitdate/>
      <!-- one repository name is required -->
      <repository/>
      <!-- Use a single langmaterial element -->
      <langmaterial>
        <!-- For each language use a separate language element
        - We can add English or any other language(s) as the default language for all the
        descriptions within an institution
        - The langcode attribute is required; we should be able to populate this
        if the language is present as defined content
```

# What is good about EAD

- Enables data exchange
- Provide enhanced descriptions back to contributors
- Provide EAD to APE and other services
- Deals well with hierarchy
- Allows for mixed content
- Suits archival finding aids - blend of data and document
- Well-defined languages – clear rules

# What is bad about EAD

- Allows for mixed content – difficult to parse
- Large and verbose – data processing inefficient
- XML is increasingly marginalised in terms of web browsers and tools
- There are fewer XML/XSLT professionals
- We want to do things that don't fit in well with EAD schema
- JSON is compact, readable, possibly more widely useful structure
- Front-end application works with the JSON

# Descriptions stored in JSON

**Archive** **Livingstone Papers**

Expand JSON

```
admin: {"process":"scotlandnatlib@233","added":1571983753000,"d... },
agents: [{"creator":{"@link":{"type":"reference"},"summary_title":"... },
aggregation: {"hasDigitalContent":false,"creator":["David Livingstone (181... },
dip: {
  accessrestrict: "
    See http://www.nls.uk/info/index.html
  },
  accruals: "
    Open.
  },
  bioghist: "
    David Livingstone (1813-1873), African missionary and explorer, educated himself while working at a cotton factory near Glasgow; attended the medical class at Anderson College and lectures at Glasgow University, 1832; entered the service of the London Missionary Society, studied medicine and science in London; embarked as a missionary for the Cape of Good Hope, 1840; made journeys into the interior, 1841, 1842, and 1843; discovered Lake Ngami, 1849, and the Zambezi in the centre of the continent, 1851; made great exploring expedition from Cape Town northwards through West-Central Africa to Loanda and back to Quillimane, 1852-6; visited England, 1856; DCL, Oxford, and FRS; published his missionary travels, and severed his connection with the London Missionary Society, 1857; consul at Quillimane, 1858-64; commanded expedition to explore Eastern and Central Africa, 1858; discovered lakes Shirwa and Nyasa, 1859; lost his wife at Shupanga, 1862; visited England, 1864; published The Zambesi and its Tributaries (1865); started on expedition to solve the question of the Nile basin, 1865; discovered Lake Bangweolo, 1868; reached Ujiji, 1869; explored the cannibal country, enduring great sufferings, and returned, almost dying, to Ujiji, where he was rescued by Stanley, 1871; reached Unyanyembe, 1872; made further explorations to discover the sources of the Nile, and died at a village in the country of Ilala; buried in Westminster Abbey, 1874.
  },
  custodhist: "
    Acquired through various deposits.
  },
  origination: "
    David Livingstone (1813-1873)
  },
  otherfindaid: "
    See http://www.nls.uk/catalogues/online/index.html and manual catalogues: http://www.nls.uk/catalogues/manual/index.html. Also, G. W. Clendennen and I. C. Cunningham, 'David Livingstone: A Catalogue of Documents', Edinburgh, 1979.
  },
}
```

Q

Add Filter

No S

DBB1

MS 3430-3518

MS 7530-8022

KRG

GMIL

DONP

# Development and innovation

- Use of JSON data
  - better for innovation e.g. Names Project
  - JSON - easier to read and map, more flexible
  - regex expressions don't work with XML very well

# Names project

```
1 public class AgentStringAnalysis
2 {
3     Map<Integer,String> blocks = new LinkedHashMap<Integer,String>();
4     Map<Integer,List<AgentDateLocation>> yearBlocks = new LinkedHashMap<Integer,List<AgentDateLocat
5     List<Integer> punctuationEnd = new ArrayList<Integer>();
6     List<Integer> punctuationStart = new ArrayList<Integer>();
7     //Map<Integer,String> yearBlockRemainders = new LinkedHashMap<Integer,String>();
8     List<Integer> f1Blocks = new ArrayList<Integer>();
9     List<Integer> bDotBlocks = new ArrayList<Integer>();
10
11
12     boolean hasAnalysedContent()
13     {
14         return yearBlocks.size()>0 || punctuationStart.size()>0 || punctuationEnd.size()>0 || f1Blocks.size()>0
15     }
16
17     @Override
18     public String toString()
19     {
20         TextStringBuilder sb = new TextStringBuilder();
21         sb.append(blocks);
22
23         if(yearBlocks.size()>0)
24         {
25             sb.appendNewLine();
26             sb.append(yearBlocks);
27         }
28         if(punctuationStart.size()>0)
29         {
30             sb.appendNewLine();
31             sb.append(punctuationStart);
32         }
33         if(punctuationEnd.size()>0)
34         {
35             sb.appendNewLine();
```

Gradually expanding set of rules - e.g. how to deal with dates.

These along with regular expressions to create the environment for the data structuring.

# Thank-you

Jane Stevenson  
Archives Hub Manager

[jane.stevenson@jisc.ac.uk](mailto:jane.stevenson@jisc.ac.uk)





ArchivesSpace

a community served by ✦ LYRASIS

# EAD and ArchivesSpace

For Quo vadis, EAD? Current use cases of EAD  
Christine Di Bella, ArchivesSpace Program Manager  
June 2, 2021

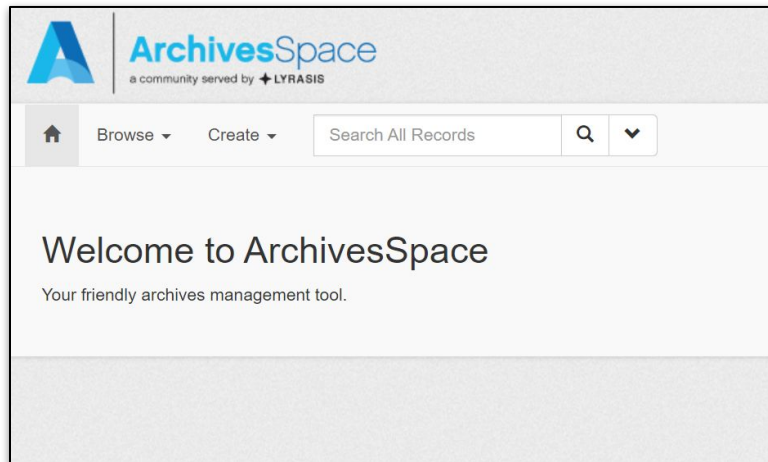


ArchivesSpace

a community served by ✦ LYRASIS



# What is ArchivesSpace?



- A community-supported, open source archival information management application that supports many archives functions, including accessioning, arrangement, description, preservation, and access.
- Used by people around the world
  - Wide variety of institutions, from small all-volunteer repositories to some of the largest research institutions in the world
  - Heaviest concentration is of U.S. users
- Ideally, a user requires zero knowledge of EAD to be able to use it, even to create EAD using it





# ArchivesSpace and standards

- First and foremost: DACS
- Imports include EAD, EAC-CPF, MARCXML (plus some CSV/spreadsheet-based importers)
- Exports include EAD, EAC-CPF, MARCXML plus DC, MODS, METS for digital objects
- Data exchange with other systems facilitated by OAI-PMH and a robust API





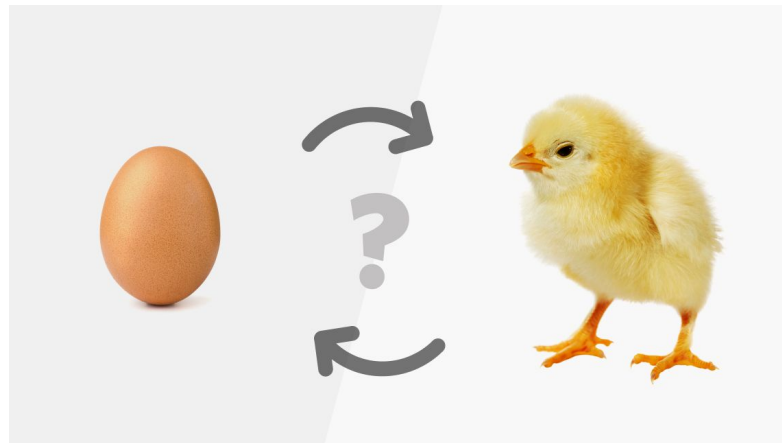
# ArchivesSpace, EAC-CPF, and EAD

- Though DACS is the basis of ArchivesSpace, its schemas are informed by the needs of some data structure standards
- Latest release made support for EAC-CPF more robust
- EAD 2002 import and export functionality included in the first release, with occasional updates in the years since
- Of note: the built-in public interface of ArchivesSpace does not rely on EAD or EAC-CPF, though people who do not use that interface or need to exchange data with another archival discovery system likely to rely on our EAD or EAC-CPF exports



# ArchivesSpace and EAD3

- EAD3 export was a community contribution in 2017
- Nearly four years later, still no EAD3 import
  - On our roadmap as a contractor project for this year
- Chicken/egg issue
  - No EAD3 import in ArchivesSpace likely means less implementation of EAD3 in U.S. repositories, but little outright demand and no community contribution means we haven't prioritized it to this point.
- Archives are sloooow
  - Hear from people all the time who would benefit from using ArchivesSpace and haven't started with systematized collection management at all
  - Puts using a 19-year old flavor of a descriptive standard in perspective





# ArchivesSpace and stating the obvious...

- EAD is problematic for systems like ArchivesSpace
  - Abundant (excessive?) flexibility
    - Even though this is very much archivally informed and guided
    - EAD3 is definitely an improvement in this area, but so far seemingly not enough that it offset needing to change procedures and supporting materials like stylesheets that people carefully built over the years
  - Description put into ArchivesSpace is used in many different ways
    - ArchivesSpace is not solely (or even primarily) an EAD authoring tool
    - Tying a system used for many purposes too closely to one data structure standard would leave people we need to support behind





# ArchivesSpace and the EAD of the future



- Zero flexibility or nuance...
  - Obviously not practical or desirable, and we don't take this approach in ArchivesSpace either
- But focus on making it easy for people who are not experts to understand, discuss and use
  - For example, with EAC-CPF in ArchivesSpace, though we added very substantially to the agents schema, we have continued to keep the requirements for creating valid agent records minimal
    - Acknowledge this means people may create records that lead to invalid EAC-CPF
    - Training and documentation to help people understand what they need to do if creating valid EAC-CPF is a goal for them
- Work closely and directly with system maintainers when revising and building migration pathways





# Thanks for listening

Feel free to reach out, with thoughts on EAD,  
standards in general, or anything else

[ArchivesSpaceHome@lyrasis.org](mailto:ArchivesSpaceHome@lyrasis.org)

[christine.dibella@lyrasis.org](mailto:christine.dibella@lyrasis.org)





# Thanks!

- The recording will be available on YouTube:

<https://www.youtube.com/user/saastaff>

- Feedback? Suggestions for ...? Please complete our survey!

<http://bit.ly/eadwebinar>